ISSN 1561-5359. Штучний інтелект, 2017, № 3-4

УДК 519.8

В.М. Горбачук, М.С. Дунаєвський, А.А. Сирку, С.-Б. Сулейманов
Інститут кібернетики імені В.М. Глушкова НАН України, Україна
пр. Глушкова, 40, м. Київ, 03187

ОПТИМІЗАЦІЙНІ ПИТАННЯ ОЦІНЮВАННЯ ЩІЛЬНОСТІ НА РЕАЛЬНИХ ДАНИХ

V.M. Gorbachuk, M.S. Dunaevskyi, A.A. Syrku, S.-B. Suleimanov
V.M. Hlushkov Institute of Cybernetics, National Academy of Sciences of Ukraine, Ukraine
40, Hlushkova av., Kyiv, 03187

THE OPTIMIZATION ISSUES OF DENSITY ESTIMATION ON REAL DATA

Для оцінювання щільності статистичного розподілу часто застосовують підхід максимальної ентропії, рівносильний підходу максимальної правдоподібності. Однак на малих наборах вхідних даних такий підхід дає надлишковість оцінки. Надлишковість оцінки можна усувати такими методами згладження як регуляризація чи переформулювання обмежень.

Ключові слова: максимізація, ентропія, вірогідність, щільність, розподіл, Больцман, Гіббс.

The maximum entropy approach, equivalent to the maximum likelihood approach, is often applied to estimation of density for a statitistical distribution. But such an approach produces estimate overfitting on small sets of input data. The estimate overfitting can be eiliminated by such smoothing techniques as regularization or reformulation of constraints.

Key words: maximization, entropy, likelihood, density, distribution, Boltzmann, Giibs.

Вступ

Для широкого класу методів максимізації ентропії та вірогідності розвивається єдиний підхід, що гарантує їхню збіжність і статистичну ефективність [1; 2]. Наприклад, можна обґрунтувати статистичну ефективність відомих методів регуляризації з цільовими функціями у стандартних нормах, а також зі спеціальними цільовими функціями, які мають кращі статистичні властивості, використовуючи інформацію про структуру простору ознак, про зсув отримання вибірки, про інші способи оцінювання щільності [3; 4]. Крім того, для розв’язання загальної задачі про максимальну ентропію (максент) можна запропонувати нові алгоритми й довести їхню збіжність, узагальнюючи методи, основані на теорії компактності, інформаційній геометрії та оптимізації [5–7].

Постановка проблеми

Пропоновані методи й алгоритми максимізації ентропії та вірогідності можна застосовувати до моделювання поширення біологічних видів [8]. Інтерес мають рідкісні види, характерні малою кількістю наявних вибірок даних, причому зсунених до місцевостей, де ці дані легше збирати, – доріг, міст, аеропортів, водних шляхів [4]. Такий зсув можна оцінювати за множиною відвіданих місцевостей. Незважаючи на малу кількість вибірок для даного виду, часто є доступ до баз даних про численні споріднені види [9; 10]. Тоді бажано використовувати методи множинного оцінювання, щоб поліпшувати якість прогнозування для всіх споріднених видів [6; 7; 10].

Наприклад, при моделюванні поширення змієїда (хижого птаха родини яструбових, який зустрічається у Канівському природному заповіднику Київського національного університету імені Тараса Шевченка і включається до Червоної книги) за допомогою пікселів на мапі Південно-Західної Євразії можна

використовувати обмеження, основане на такій ознаці, як щорічні (атмосферні) опади: обсяг середніх щорічних опадів, що сприяє поширенню змієїда, має рівнятися середнім спостережуваним опадам.

Аналіз останніх досліджень і публікацій

Принцип максимальної ентропії походить від статистичної механіки. Запропонований у статистичній механіці підхід оцінювання щільності [11–13] поширився на інші галузі, зокрема, на комп'ютерну обробку природної мови [14; 15]. Вважається, що для шуканої щільності розподілу задано набір відомих обмежень, які часто основані на множині вибірок шуканого розподілу і виражені через множину ознак (функцій, значеннями яких є дійсні числа), визначених на просторі вибірок [16]. Як правило, обмеження вимагають відповідності сподіваної ознаки та її емпіричного середнього.

Мета дослідження

Висвітлити оптимізаційні питання, що виникають при машинному моделюванні поширення видів, тобто оцінюванні щільності виду за пікселями на мапі. Ознаками є прості функції змінних довкілля, а обмеження виходять із частоти спостережень видів.

Виклад основного матеріалу

Для визначення розподілу, що має максимальну ентропію, можна застосувати метод множників Лагранжа. Відповідно до умов Каруша–Куна–Такера (Karush–Kuhn–Tucker) цей розподіл є розподілом з максимальною вірогідністю з класу експоненційних розподілів, де ознаки відіграють роль достатніх статистик. У статистичній механіці такі розподіли називають розподілами Гіббса (Gibbs).

Вивчаючи властивості газів як систем, що складаються з великої кількості молекул, Людвіг Больцман (1844–1906, Австро-Угорщина) також відповідав на фундаментальне питання залежності макроскопічного стану (макростану) системи від її мікроскопічних властивостей. Макростан включає такі властивості системи як її об'єм, загальне число молекул, загальну енергію, а мікростан – такі властивості окремих молекул системи як їхні швидкості та положення.

Для простоти Больцман припустив, що молекули газу змінюють свій стан (свої координати) k дискретно, тобто їхні швидкості та положення є дискретними, набуваючи значення $1, 2, ..., K$. Енергія – це ключова характеристика як макростану, так і мікростану. Енергія кожної молекули – це сума кінетичної енергії, що залежить лише від швидкості молекули, та потенційної енергії, що залежить лише від положення молекули в силовому полі. Нехай дискретні стани є малими настільки, що молекули однакового стану k мають (майже) однакову енергію E_k, але водночас ці стани є великими настільки, що багато молекул має однаковий стан. Тоді мікростан системи задається вектором, елементи якого відповідають станам усіх її молекул, а макростан визначається гістограмою числа N_k молекул за кожним станом k. Таким чином, для визначення макростану достатньо обчислювати найвірогіднішу гістограму.

Застосовуючи принцип байдужості, Больцман вважав, що всі мікростани є рівноймовірними. Тому найвірогіднішу гістограму можна отримати за найбільшим числом мікростанів. Якщо загальне число молекул становить $N = \sum_{k=1}^{K} N_k$, то загальне число способів їхнього розподілу за станами визначається мультиноміальним коефіцієнтом

$$c(N_1, N_2, ..., N_K) = \frac{N!}{N_1! N_2! ... N_K!}, \quad (1)$$

який слід максимізувати при законі збереження загальної енергії системи

$$E = \sum_{k=1}^{K} N_k E_k . \quad (2)$$

Для максимізації функції (1) зручніше користуватися її логарифмом, який при множенні на константу Больцмана дає термодинамічну ентропію. Для цього логарифма скористаємося наближенням Стірлінга (Stirling)

$$\ln\left(\frac{N!}{N_1! N_2! ... N_K!}\right) \approx \left(\sum_{k=1}^{K} N_k \ln \frac{N}{N_k}\right).$$

Звернімо увагу, що $\frac{N_k}{N} = p_k$ – це частота стану k. Тоді задача Больцмана максимізації функції (1) при обмеженні (2) зводиться до задачі максимізації по $p_1, p_2, ..., p_K$ функції термодинамічної ентропії

$$f(p_1, p_2, ..., p_K) = \left(\sum_{k=1}^{K} N_k \ln \frac{1}{p_k}\right) \quad (3)$$

при обмеженні для середньої енергії молекули

$$\frac{E}{N} = \sum_{k=1}^{K} \frac{N_k}{N} E_k = \sum_{k=1}^{K} p_k E_k . \quad (4)$$

Звідси, використовуючи метод множників Лагранжа, знаходимо

$$p_k \approx e^{\lambda E_k},$$

де λ – множник Лагранжа, що відповідає обмеженню (4). Знайдений вираз для розподілу Больцмана може стати основою для вивчення різних властивостей газів, наприклад, розподілу щільності газу у гравітаційному полі.

Інтерпретуючи методологію Больцмана на мові теорії інформації [17], можна розв’язувати загальніші проблеми: «статистична механіка може стати лише окремим прикладом статистичних рішень» [11; 12]. Замінюючи термодинамічну ентропію на теоретико-інформаційну ентропію, можна вимірювати нашу невизначеність щодо даної системи. Коли наші знання щодо даної системи виражаються обмеженням (4), то серед усіх розподілів слід обрати найбільш незалежний від пропущеної в цьому обмеженні інформації – розподіл з найбільшою теоретико-інформаційною ентропією

$$H(p) = -\sum_{k=1}^{K} p_k \ln p_k ,$$

пов’язану з функцією (3). Розподілу Больцмана задовольняє гранична межа перехідної ймовірності у випадковому локальному пошуку [18].

Принцип максимальної ентропії можна вважати узагальненням принципу байдужості. У статистичних рішеннях цей принцип дозволяє знаходити розподіл, який максимізує ентропію при даних обмеженнях. У задачі Больцмана єдина ознака – це енергія, а розподіл Больцмана – це випадок розподілу Гіббса, виокремлений однією ознакою. Якщо x – довільний стан у просторі X, $\vec{f}(x)$ – вектор значень ознак, які характеризують цей стан, то результуючий розподіл Гіббса має вид

$$p(x) \approx e^{\vec{\lambda} \vec{f}(x)},$$

де $\vec{\lambda}$ – вектор множників Лагранжа, що відповідають обмеженням.

Теоретико-інформаційне обґрунтування [11; 12] можна узагальнювати, припускаючи наявність певного (початкового) розподілу q_0 за відсутності будь-яких даних [19; 20]. Тоді можна обирати найближчий до q_0 розподіл серед усіх розподілів, які задовольняють заданим обмеженням. Міра близькості – це відносна ентропія

$$D(p \| q_0) = \sum_{x \in X} p(x) \ln \frac{p(x)}{q_0(x)},$$

яку називають дивергенцією (divergence) Кульбака–Лейблера (Kullback–Leibler); вона вимірює обсяг інформації про шуканий результат, який можна дістати при знанні p замість q_0. Якщо q_0 – рівномірний розподіл, то критерій мінімальної відносної ентропії замінюватиметься критерієм максимальної ентропії, а результуючі розподіли Гіббса матимуть вигляд

$$p(x) \approx q_0(x) e^{\vec{\lambda}\, \vec{f}(x)}.$$

Слід пояснити, як застосовувати до задачі оцінювання щільності такі теоретико-інформаційні величини, як ентропія чи відносна ентропія. Хоча багато досліджень даної задачі має теоретико-інформаційне обґрунтування [21], є інші теоретичні обгрунтування цієї задачі – теорія великих відхилень, аксіоматична теорія, теорія ігор.

Теорія великих відхилень вивчає ймовірності маловірогідних подій, які у статистичній механіці відповідають макростанам (гістограмам) з ентропією, нижчою за максимальну. Наприклад, у задачі Больцмана число реалізацій емпіричного розподілу p для N частинок становить

$$c(N_1, N_2, ..., N_K) = \frac{N!}{N_1! N_2! ... N_K!} = e^{N(H(p)+o(1))}, \qquad (5)$$

де $o(1) \to 0$ при $N \to \infty$, $H(p)$ – ентропія розподілу p. Отже, розподіл $\hat{p}$ з максимальною ентропією реалізується $e^{N(H(\hat{p})+o(1))}$ макростанами. Порівняємо це число і загальне число реалізацій з ентропією, нижчою за максимальну, тобто з ентропією, нижчою за $H(\hat{p}) - \varepsilon$, де $0 < \varepsilon$ – задана величина.

Якщо $H(p) \le H(\hat{p}) - \varepsilon$, то в силу залежності (5) число реалізацій довільної гістограми p не перевищує $e^{N(H(\hat{p})-\varepsilon+o(1))}$. Число таких гістограм обмежене зверху загальним числом $(N+1)^K$ гістограм для N частинок, кожна з яких перебуває в одному з K станів. Отже, загальне число реалізацій з ентропією, нижчою за $H(\hat{p}) - \varepsilon$, не перевищує величини $(N+1)^K\, e^{N(H(\hat{p})-\varepsilon+o(1))}$, яка при досить великих значеннях N експоненційно менша числа $e^{N(H(\hat{p})+o(1))}$ реалізацій роподілу $\hat{p}$. Тому серед усіх емпіричних розподілів, що задовольняють обмеженням, лише експоненційно мала частка не потрапляє у довільно малий окіл розподілу з максимальною ентропією.

Вищезазначені міркування припускають, що всі реалізації розподілу є рівноймовірними a priori. Коли кожна молекула перебуває у своєму стані відповідо до деякого апріорного розподілу q_0, то ентропію $H(p)$ слід замінювати від'ємною відносною ентропією $-D(p \| q_0)$.

Такі міркування узагальнює теорема Санова про те, що емпіричний розподіл при обмеженнях наближує розподіл з максимальною ентропією [22]. Цю теорему суттєво узагальнюють умовні граничні теореми [23–25]. Наприклад, коли обмеження задають опуклу множину розподілів ймовірності, а вибірки належать незалежним реалізаціям розподілу q_0, яке не задовольняє даним обмеженням, то умовний розподіл першої вибірки збігається до розподілу $\hat{p}$ з максимальною ентропією при умові, що емпіричний розподіл p задовольняє цим обмеженням [24]. Якщо теорема Санова стосується емпіричного розподілу для N частинок, то теорема Чіжара стосується граничного розподілу для єдиної частинки.

Теорема Санова й умовні граничні теореми характеризують властивості емпіричного стану системи при відомих умовах емпіричного стану системи за відомих умов і припущення, що цей стан генерується розподілом q_0. Хоча ці теореми суворо обґрунтовують максент у статистичній механіці [26], вони потребують розвитку у статистичних рішеннях для пошуку невідомого генеруючого розподілу з його початковою оцінкою q_0.

В аксіоматичних підходах до статистичних рішень висуваються такі вимоги (властивості) для консистентного статистичного рішення, як інваріантність при змінах координат і збереження консистентності при розкладах системи на окремі підсистеми [21; 27; 28]. Доведено, що єдиний метод статистичних рішень, який задовольняє водночас всім цим вимогам, – це принцип максимальної ентропії.

Проте не всі зазначені вимоги є самоочевидними: вимога мінімальних знань [11; 12], вимога найменшої відносної ентропії [19], вимоги властивостей консистентності [21] не пов’язані безпосередньо з оцінюванням щільності [29].

Якість рішення часто оцінюється на контрольній множині, що складається з вибірок, відсутніх під час навчання. Частоти, спостережувані в контрольній множині, як правило, вважаються наближеннями деяких граничних ймовірностей, отримуваних при нескінченно великій кількості вибірок. Такі ймовірності вважаються істинними при частотній інтерпретації [30].

У машинному навчанні та статистиці найпоширенішою альтернативою частотній інтерпретації є Байєсова інтерпретація, за якою існують апріорні ймовірності для всіх щільностей у даному класі. Після спостереження класична Байєсова оцінка дає апостеріорний розподіл для всіх можливих щільностей, оснований на апріорному розподілі та цьому спостереженні; некласична Байєсова оцінка дає єдину щільність, екстремальну для апостеріорного розподілу.

Якщо частотна і Байєсова інтерпретації пов’язують задачі оцінювання щільності із спостережуваними вибірками безпосередньо (через частоти) чи опосередковано (через апостеріорний розподіл), то для принципу максимальної ентропії такий зв’язок потребує вивчення. Підхід теорії інформації, оснований на припущенні максимальної байдужості, дає класичну інтерпретацію ймовірностей [30] – інтерпретацію, яку застосовують для аналізу перестановок у колоді карт. Подібно до припущення Больцмана про рівноймовірність кожного макростану, кожна перестановка теж вважається рівноймовірною. В аксіоматичних підходах принцип байдужості замінюється множиною вимог консистентності, залишаючи відкритими питання отримання обмежень зі спостережень і генерування вибірок.
Теоретико-ігровий підхід до обґрунтування принципу максимальної ентропії [31] пов’язує генерування вибірок з оцінюванням щільності.

Максент на мові теорії ігор нагадує класичне прийняття статистичних рішень – рішень при оцінюванні щільності, коли один гравець (природа) обирає довільний розподіл π, який задовольняє відомій (обом гравцям) множині обмежень, а інший гравець (особа, яка приймає рішення (ОПР)), не знаючи π, обирає такий розподіл q, який максимізує логарифм вірогідності відносно логарифма вірогідності при початковому розподілі q_0, задовольняючи множину обмежень. Отже, ОПР, намагаючись максимізувати свою цільову функцію

$$f(\pi(x), q(x)) = \sum_{x \in X} \pi(x) \ln q(x) - \sum_{x \in X} \pi(x) \ln q_0(x) = \sum_{x \in X} \pi(x) \ln \frac{q(x)}{q_0(x)},$$

обирає стратегію максиміну

$$\tilde{q} = \arg\max_{q \in Q} \min_{\pi \in \Pi} f(\pi(x), q(x)), \tag{6}$$

де Q – множина всіх щільностей на даному просторі вибірок, Π – множина щільностей, які задовольняють даним обмеженням. Показано, що ця стратегія рівносильна щільності з мінімальною відносною ентропією [19] (щільності з максимальною ентропією [11; 12])

$$\hat{p} = \arg\min_{p \in \Pi} D(p \,\|\, q_0).$$

Замість стратегії максиміну можна застосовувати стратегію максимального сподіваного значення

$$\tilde{q} = \arg\max_{q \in Q} E_{\pi \in \Pi}[f(\pi(x), q(x))], \tag{7}$$

де $E_{\pi \in \Pi}$ означає сподіване (expected) значення за $\pi \in \Pi$ [10].

Максимізацію вірогідності застосовують в оптимальному кодуванні й оптимальному виборі ігрових ставок [32]. Якщо при застосуванні максимізації ентропії у теоретико-інформаційному та теоретико-іровому підходах висуваються бажані вимоги до шуканого розподілу, то при застосуванні максимізації вірогідності безпосередньо визначається критерій оптимальності відносно розподілу, який вважається істинним. Оскільки припущення про існування деякого єдиного істинного розподілу (який може бути граничним для нескінченно великої кількості вибірок) основано на понятті частоти, то стратегію максиміну при максимізації вірогідності можна вважати частотною інтерпретацією максимізації ентропії.

За умовами Каруша–Куна–Такера розподіл, що має максимальну ентропію при обмеженнях-рівностях для емпіричних середніх, є розподілом з максимальною вірогідністю, який належить класу експоненційних розподілів. Отже, максимізація ентропії має альтернативну інтерпретацію як максимізація вірогідності [33; 34]. Однак максимізація вірогідності у класичній статистиці [35] відрізняється від максимізації ентропії: при максимізації вірогідності істинний розподіл припускається належним до того класу розподілів, за якими максимізується вірогідність, а при максимізації ентропії подібного параметричного припущення немає; максимізація вірогідності є швидше оцінюванням параметрів з вивченням асимптотичних властивостей їхніх оцінок, ніж оцінюванням щільності, а максимізація ентропії є порівнянням швидше фактичних розподілів (розподілу з максимальною ентропією з істинним розподілом або його найкращим наближенням через розподіл Гіббса), ніж параметрів.

Важливе питання для будь-якого застосування принципу максимальної ентропії – це вибір множини обмежень. Хоча найпоширенішими є обмеження-рівності [21; 22; 29], відомі інші типи обмежень [11; 12; 21; 36; 37]. При пошуку розв'язку задачі (6) суттєво, щоб істинний розподіл належав множині Π, яка не повинна бути занадто великою. З метою належного врахування множини Π пропонується критерій (7) [38].

Коли обмеження-рівності основані на емпіричних середніх [39], то максент може вести до надлишковості навчальних даних. Водночас для кожної змінної довкілля характерні порогові ознаки, які набувають бінарних значень (значення 1, якщо змінна довкілля перевищує заданий поріг, і значення 0 в інших випадках). Коли кожна така змінна має безліч ознак, то розподіл з максимальною ентропією буде нетривіальним лише на значеннях, досягнутих вибірками.

Висновки

Загалом проблема максимізації вірогідності полягає у тому, що емпіричні середні ознак майже завжди не дорівнюватимуть їхнім істинним очікуванням, а тому цільовий розподіл не задовольнятиме обмеженням, які накладаються на результуюче обмеження. Крім того, проблема ускладнюється малими розмірами вибірок. При інтерпретації максимізації ентропії як максимізації вірогідності та звуженні шуканих розподілів до класу експоненційних розподілів можлива надлишковість оцінок.

Література

1. Горбачук В.М. Статистическое распознавание образов и стохастическая оптимизация / Распознавание образов и анализ изображений: новые информационные технологии. – Минск: ИТК АН БССР, 1991. – С. 74–76.
2. Горбачук В. Інформатика як інтегральна наука // Інформація та нові технології. – 1993. – №. 1. – С. 11–16.
3. Fischer G., Ermolieva T., Ermoliev Y., van Velthuizen H.T. Sequential downscaling methods for estimation from aggregate data. – Laxenburg, Austria: IIASA, 2006. – IIASA Interim Report IR-06-002. – 12 p.
4. Atoyev K.L., Golodnikov O.M., Gorbachuk V.M., Ermoliev Yu.M., Ermolieva T.Yu., Kiriljuk V.S., Knopov P.S. The mathematical problems of complex management and effective utilization of food, energy and water resources under increased uncertainties / Integrated management, security, and robustness. A.G. Zagorodny, Yu.M. Ermoliev, V.L. Bogdanov (eds.) – Kyiv: NAS of Ukraine, 2014. – P. 198–227.
5. Гайворонский А.А., Горбачук В.М. Интерактивная система для решения задач / Моделирование плановых расчетов и диалоговая оптмизация. – К.: Знание, 1990. – С. 17–18.
6. Горбачук В. Економетричне програмування TSP та EViews. – Препр. 96-14. – К.: Ін-т кібернетики НАН України, 1996. – 24 с.
7. Chikrii A., Denisova N., Gorbachuk V., Gromaszek K., Krivonos Y., Lytvynenko V., Matychyn I., Osypenko V., Smailova S., Wojcik W. Current problems in information and computational technologies. V. 2. W. Wojcik, J. Sikora (eds.) – Lublin: Politechnika Lubelska, 2012. – 196 p.
8. Горбачук В.М., Кирилюк О.П. Моделирование рационального рыбного промысла в водоемах // Математическое моделирование в проблемах рационального природопользования. – Ростов-на-Дону: РГУ, 1991. – С. 11–12.
9. Горбачук В.М., Дунаєвський М.С. Аналіз і застосування прикладної загальної рівноваги // Наукові записки НаУКМА. – 2010. – Т. 107. – С. 96–100.
10. Горбачук В.М., Єрмольєв Ю.М., Єрмольєва Т.Ю. Двоетапна модель еколого-економічних рішень // Вісник Одеського національного університету. Економіка. – 2016. – Т. 21. – Вип. 9. – С. 142–147.
11. Jaynes E.T. Information theory and statistical mechanics // Physical review. – 1957. – 106(4). – P. 620–630.
12. Jaynes E.T. Information theory and statistical mechanics / Jaynes E.T. Papers on probability, statistics, and statistical physics. R.D.Rosenkrantz (ed.) – Dordrecht, Holland: D.Reidel Publishing Company, 1983. – P. 6–16.
13. Kapur J.N., Kesavan H.K. Entropy optimization principles with applications. – Academic Press, 1992.
14. Berger A.L., Della Pietra S.A., Della Pietra V.J. A maximum entropy approach to natural language processing // Computational linguistics. – 1996. – 22(1). – P. 39–71.
15. Della Pietra S., Della Pietra V., Lafferty J. Inducing features of random fields // IEEE transactions on pattern analysis and machine intelligence. – 1997. – 19(4). – P. 1–13.

16. Горбачук В.М., Кошулько А.І., Сирку А.А. До природних обмежень економічної діяльності / Розвиток сучасних міжнародних економічних відносин: фінансово-економічні та соціальні чинники (23–24 вересня 2016 р., Одеса). – Одеса: ОНУ імені І.І. Мечникова, 2016. – С. 125–128.
17. Shannon C.E. A mathematical theory of communication // The Bell system technical journal. – 1948. – 27. – P. 379–423, 623–656.
18. Сергиенко И.В., Шило В.П. Задачи дискретной оптимизации. Проблемы, методы решения, исследования. – К.: Наук.думка, 2003. – 261 с.
19. Kullback S. Information theory and statistics. – New York, NY: Wiley, 1959.
20. Benes V.E. Mathematical theory of connecting networks and telephone traffic. – New York, NY: Academic Press, 1965.
21. Shore J.E., Johnson R.W. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy // IEEE transactions on information theory. – 1980. – 26(1). – P. 26–37.
22. Санов И.Н. О вероятности больших отклонений случайных величин // Математический сборник. – 1957. – 42(84). – № 1. – С. 11–44.
23. Van Campenhout J.M., Cover T.M. Maximum entropy and conditional probability // IEEE transactions on information theory. – 1981. – 27. – P. 483–489.
24. Csiszar I. Sanov property, generalized I-projection and a conditional limit theorem // Annals of probability. – 1984. – 12(3). – P. 768–793.
25. Grünwald P. Strong entropy concentration, game theory, and algorithmic randomness / COLT/EuroCOLT 2001: Proceedings of the 14-th Annual Conference on Computational Learning Theory and 5-th European Conference on Computational Learning Theory. – London: Springer-Verlag, 2001. – P. 320–336.
26. Csiszar I. Maxent, mathematics, and information theory / Proceedings of the 15-th Internationl Workshop on Maximum Entropy and Bayesian Methods. – Dordrecht: Kluwer Academic Publishers, 1995.
27. Skilling J. The axioms of maximum entropy / Maximum-entropy and Bayesian methods in science and engineering. Volume 1. G.J. Erickson, C.R. Smith (eds.). – Kluwer Academic Publishers, 1988. – P. 173–187.
28. Csiszar I. Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems // Annals of statistics. – 1991. – 19(4). – P. 2032–2066.
29. Dudik M. Maximum entropy density estimation and modeling geographic distributions of species. PhD (Computer Science) Dissertation. – Proнceton, NJ: Princeton University, 2007. – 179 p.
30. Hajek A. Interpretations of probability / The Stanford encyclopedia of philosophy. E.N.Zalta (ed.) – http://plato.stanford.edu/
31. Topsoe F. Information theoretical optimization techniques // Kybernetika. – 1979. – 15(1). – P. 8–27.
32. Cover T.M., Thomas J.A. Elements of information theory. – Wiley, 1991.
33. Jaynes E.T. Where do we stand on maximum entropy? / The maximum entropy formalism. R.D. Levine, M.Tribus (eds.) – Cambridge, MA: MIT Press, 1978. – P. 15–118.
34. Jaynes E.T. Where do we stand on maximum entropy? / Jaynes E.T. Papers on probability, statistics, and statistical physics. R.D. Rosenkrantz (ed.) – Dordrecht, Holland: D. Reidel Publishing Company, 1983. – P. 210–314.
35. Lehmann E.L., Casella G. Theory of point estimation. 2-nd ed. – New York, NY: Springer-Verlag, 1998.
36. Csiszar I. I-divergence geometry of probability distributions and minimization problems // Annals of probability. – 1975. – 3(1). – P. 146–158.
37. Khudanpur S.P. A method of maximum entropy estimation with relaxed constraints / Proceedings of the Johns Hopkins University Language Modeling Workshop, 1995. – P. 1–17.
38. Горбачук В.М., Сирку А.А., Сулейманов С.-Б. Моделі аналізу охоплення нестандартних даних // Компьютерная математика. – 2017. – № 1. – С. 63–72.
39. Knopov P.S., Kasitskaya E.J. Empirical estimates in stochastic optimization and identification. – Springer Science & Business Media, 2002. – 250 p.

Literatura

1. Gorbachuk V.M. Statisticheskoye raspoznavanie obrazov i stokhasticheskaya optimizatsiya / Raspoznavanie obrazov i analiz izobrazhenij: novye informatsionnye tekhologii. – Minsk: ITK AN BSSR, 1991. – S. 74–76.
2. Gorbachuk V. Informatyka yak integral'na nauka // Informatsiya ta novi tekhnologii. – 1993. – №. 1. – S. 11–16.
3. Fischer G., Ermolieva T., Ermoliev Y., van Velthuizen H.T. Sequential downscaling methods for estimation from aggregate data. – Laxenburg, Austria: IIASA, 2006. – IIASA Interim Report IR-06-002. – 12 p.
4. Atoyev K.L., Golodnikov O.M., Gorbachuk V.M., Ermoliev Yu.M., Ermolieva T.Yu., Kiriljuk V.S., Knopov P.S. The mathematical problems of complex management and effective utilization of food, energy

and water resources under increased uncertainties / Integrated management, security, and robustness. A.G .Zagorodny, Yu.M. Ermoliev, V.L. Bogdanov (eds.) – Kyiv: NAS of Ukraine, 2014. – P. 198–227.
5. Gaivoronski A.A., Gorabchuk V.M. Interaktivnaya sistema dlya resheniya zadach / Modelirovanie planovykh raschetov i dialogovaya optimizatsiya. – K.: Znanie, 1990. – S. 17–18.
6. Gorbachuk V. Ekonometrychne programuvannya TSP ta Eviews. – Prepr. 96-14. – K.: Institut kibernetiki NAN Ukrainy, 1996. – 24 s.
7. Chikrii A., Denisova N., Gorbachuk V., Gromaszek K., Krivonos Y., Lytvynenko V., Matychyn I., Osypenko V., Smailova S., Wojcik W. Current problems in information and computational technologies. V. 2. W. Wojcik, J. Sikora (eds.) – Lublin: Politechnika Lubelska, 2012. – 196 p.
8. Gorbachuk V.M., Kyrylyuk O.P. Modelirovanie ratsional’nogo rybnogo promysla v vodoemakh // Matematicheskoe modelirovanie v problemakh ratsional’nogo propodopol’zovania. – Rostov-na-Donu: RGU, 1991. – S. 11–12.
9. Gorbachuk V.M., Dunaevskyi M.S. Analiz i zastosuvannya prykladnoi zagal’noi rivnovagy // Naukovi zapysky NaUKMA. – 2010. – T. 107. – S. 96–100.
10. Gorbachuk V.M., Ermoliev Y.M., Ermolieva T.Y. Dvoetapna model’ ekologo-ekologichnykh rishen’ // Visnyk Odes’kogo natsional’nogo universytetu. Ekonomika. – 2016. – T. 21. – Вип. 9. – С. 142–147.
11. Jaynes E.T. Information theory and statistical mechanics // Physical review. – 1957. – 106(4). – P. 620–630.
12. Jaynes E.T. Information theory and statistical mechanics / Jaynes E.T. Papers on probability, statistics, and statistical physics. R.D. Rosenkrantz (ed.) – Dordrecht, Holland: D. Reidel Publishing Company, 1983. – P. 6–16.
13. Kapur J.N., Kesavan H.K. Entropy optimization principles with applications. – Academic Press, 1992.
14. Berger A.L., Della Pietra S.A., Della Pietra V.J. A maximum entropy approach to natural language processing // Computational linguistics. – 1996. – 22(1). – P. 39–71.
15. Della Pietra S., Della Pietra V., Lafferty J. Inducing features of random fields // IEEE transactions on pattern analysis and machine intelligence. – 1997. – 19(4). – P. 1–13.
16. Gorbachuk V.M., Koshul’ko A.I., Syrku A.A. Do pryrodnykh obmezhen’ ekonomichnoi diayal’nosti / Rozvytok suchasnykh mizhanorodnykh ekonomichnykh vidnosyn: finansovo-eknomichni ta sotsial’ni chynnyky (23–24 veresnya 2016 r., Odesa). – Odesa: ONU imeni I.I. Mechnikova, 2016. – S. 125–128.
17. Shannon C.E. A mathematical theory of communication // The Bell system technical journal. – 1948. – 27. – P. 379–423, 623–656.
18. Sergienko I.V., Shylo V.P. Zadachi diskretnoj optimizatsii. Problemy, metody resheniya, issledovaniya. – K.: Nauk.dumka, 2003. – 261 s.
19. Kullback S. Information theory and statistics. – New York, NY: Wiley, 1959.
20. Benes V.E. Mathematical theory of connecting networks and telephone traffic. – New York, NY: Academic Press, 1965.
21. Shore J.E., Johnson R.W. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy // IEEE transactions on information theory. – 1980. – 26(1). – P. 26–37.
22. Sanov I.N. O veroyatnosti bol’shikh otklonenij sluchajnykh velichin // Matematicheskij sbornik. – 1957. – 42(84). – № 1. – С. 11–44.
23. Van Campenhout J.M., Cover T.M. Maximum entropy and conditional probability // IEEE transactions on information theory. – 1981. – 27. – P. 483–489.
24. Csiszar I. Sanov property, generalized I-projection and a conditional limit theorem // Annals of probability. – 1984. – 12(3). – P. 768–793.
25. Grünwald P. Strong entropy concentration, game theory, and algorithmic randomness / COLT/EuroCOLT 2001: Proceedings of the 14-th Annual Conference on Computational Learning Theory and 5-th European Conference on Computational Learning Theory. – London: Springer-Verlag, 2001. – P. 320–336.
26. Csiszar I. Maxent, mathematics, and information theory / Proceedings of the 15-th Internationl Workshop on Maximum Entropy and Bayesian Methods. – Dordrecht: Kluwer Academic Publishers, 1995.
27. Skilling J. The axioms of maximum entropy / Maximum-entropy and Bayesian methods in science and engineering. Volume 1. G.J. Erickson, C.R.Smith (eds.). – Kluwer Academic Publishers, 1988. – P. 173–187.
28. Csiszar I. Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems // Annals of statistics. – 1991. – 19(4). – P. 2032–2066.
29. Dudik M. Maximum entropy density estimation and modeling geographic distributions of species. PhD (Computer Science) Dissertation. – Pronceton, NJ: Princeton University, 2007. – 179 p.
30. Hajek A. Interpretations of probability / The Stanford encyclopedia of philosophy. E.N.Zalta (ed.) – http://plato.stanford.edu/
31. Topsoe F. Information theoretical optimization techniques // Kybernetika. – 1979. – 15(1). – P. 8–27.
32. Cover T.M., Thomas J.A. Elements of information theory. – Wiley, 1991.

33. Jaynes E.T. Where do we stand on maximum entropy? / The maximum entropy formalism. R.D.Levine, M.Tribus (eds.) – Cambridge, MA: MIT Press, 1978. – P. 15–118.
34. Jaynes E.T. Where do we stand on maximum entropy? / Jaynes E.T. Papers on probability, statistics, and statistical physics. R.D. Rosenkrantz (ed.) – Dordrecht, Holland: D.Reidel Publishing Company, 1983. – P. 210–314.
35. Lehmann E.L., Casella G. Theory of point estimation. 2-nd ed. – New York, NY: Springer-Verlag, 1998.
36. Csiszar I. I-divergence geometry of probability distributions and minimization problems // Annals of probability. – 1975. – 3(1). – P. 146–158.
37. Khudanpur S.P. A method of maximum entropy estimation with relaxed constraints / Proceedings of the Johns Hopkins University Language Modeling Workshop, 1995. – P. 1–17.
38. Gorbachuk V.M., Syrku A.A., Suleimanov S.-B. Modeli analizu okhoplennya nestandartnykh danykh // Kompyuternaya matematika. – 2017. – № 1. – C. 63–72.
39. Knopov P.S., Kasitskaya E.J. Empirical estimates in stochastic optimization and identification. – Springer Science & Business Media, 2002. – 250 p.

RESUME

V.M. Gorbachuk, M.S. Dunaevskyi, A.A. Syrku, S.-B. Suleimanov
The optimization issues of density estimation on real data

Density estimation remains to be the major problem of statistics and statistical learning theory. In practice, a few samples are observed, samples are biased, and a target density is related with other densities. This indicates to the importance of efficient utilization of training data. For example, at a small number of samples in multidimensional space, one should use the information for all the dimensions while many classical methods don’t do it. Thus, the methods using a large number of dimensions for small data sets are needed. If the data are collected with a bias, then any knowledge about that bias would improve the efficiency of prediction.

When several densities are estimated on overlapped or ordered sets of data (signals), then the signal joint for those densities would improve the accuracy of individual estimates. Thus, the multiple estimation techniques balancing individual information about a set of data and the joint information about sets of data.

While many authors have studies the issues on maximization of entropy (or application of Gibbs distributions), the issues on choice of constraints appropriate for a target distribution has needed a new research.

By the Karush–Kuhn–Tucker conditions, a distribution having the maximal entropy under equality constraints for empirical averages is the distribution with maximal likelihood belonging to a class of exponential distributions.

Thus, an entropy maximization has the alternate interpretaion as a likelihood maximization. But the likelihood maximization in classical statistics is different from an entropy maximization: at the likelihood maximization, a true distribution is assumed to belong to the class of distributions where the likelihood is maximized, whereas at an entropy maximization a similar parametric assumption is absent; a likelihood maximization is rather parameter estimation with study of asymptotic properties for estimates than density estimation, while an entropy maximization is the comparison of rather actual distributions than parameters.

Надійшла до редакції 03.11.2017